DOCUMENT RESUME

ED 470 203 TM 034 343

AUTHOR Kester, Ellen Stubbe; Pena, Elizabeth D.

TITLE Limitations of Current Language Testing Practices for Bilinguals. ERIC Digest.

INSTITUTION ERIC Clearinghouse on Assessment and Evaluation, College Park, MD.

SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.

REPORT NO EDO-TM-02-03

PUB DATE 2002-07-00

NOTE 4p.

CONTRACT ED-99-CO-0032

AVAILABLE FROM ERIC Clearinghouse on Assessment and Evaluation, 1129 Shriver Laboratory, University of Maryland, College Park, MD 20742. Tel: 800-464-3742 (Toll Free); Web site: http://ericae.net.

PUB TYPE ERIC Publications (071) -- ERIC Digests in Full Text (073)

EDRS PRICE EDRS Price MF01/PC01 Plus Postage.

DESCRIPTORS Bilingual Education; *Bilingual Students; Bilingualism; Elementary Secondary Education; Higher Education; *Language Tests; *Psychometrics; *Test Use; *Testing Problems; *Translation

IDENTIFIERS ERIC Digests

ABSTRACT

Current practices for assessment of language in bilingual students frequently involve the use of tests translated from English to the target language or tests designed for, and normed on, monolingual people. This Digest explains why these common approaches are not well suited for a bilingual population and provides guidance regarding more suitable approaches. There are many problems with test translation. Tests translated from one language to another do not retain their psychometric properties, and issues of equivalence of both vocabulary and grammar surface. Bilingual school children usually are circumstantial bilinguals, and their vocabularies in the different languages may reflect the different circumstances in which the vocabulary is used. Language tests can be improved if test developers: (1) ensure that concepts and linguistic features are appropriately represented for each language; (2) use conceptual scoring systems to eliminate underestimation of ability; (3) select an appropriate mix of item types to gain the maximal amount of information about language ability in each language; and (4) consider the frequency of occurrence of the words. (Contains 11 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made from the original document.

ED 470 203

Limitations of Current Language Testing Practices for Bilinguals

Ellen Stubbe Kester and *Elizabeth D. Peña*
The University of Texas at Austin

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM034343

ERIC Digest EDO-TM-02-03 July 2002

BEST COPY AVAILABLE

Limitations of Current Language Testing Practices for Bilinguals

Ellen Stubbe Kester and *Elizabeth D. Peña*
The University of Texas at Austin

Few diagnostic tools are designed explicitly for children who are exposed to two languages (Valdés & Figueroa, 1994). Current practices for assessment of language in bilinguals frequently involve the use of tests that are translated from English to the target language and/or tests designed for and normed on monolinguals. This Digest explains why these common approaches are not well suited for a bilingual population and provides guidance to test developers and administrators regarding more suitable approaches.

Problems with Test Translation

When tests are translated from one language to another, they do not retain their psychometric properties. Of particular interest in the assessment of language is the developmental order in which target features of the language are learned. Translating a test from one language to another -- typically from English -- may mean that items are organized by order of English difficulty, rather than reflecting the developmental order of the target language. The translated Spanish version of the Preschool Language Scale-3 provides an illustration. Restrepo and Silverman (2001) found several item difficulty discrepancies between the original English and the translated Spanish version when tested with predominately Spanish-speaking preschoolers. For example, items related to prepositions, which were relatively easy for English speakers, were more difficult for Spanish speakers. On the other hand, the "function" items requiring students to point out objects based on a description of their use (something like "Show me what people use for cooking" or "What do you sweep with?") were easier for the Spanish speakers than the English speakers.

Figueroa (1989) noted that words may generally represent the same concept but have variations and different levels of difficulty across languages. An illustration of this is found in a study of vocabulary test translations (Tamayo, 1987). When test items were translated from English to Spanish, they differed in frequency of occurrence in each language. Because the Spanish translations were of lower frequency within Spanish, test scores obtained from Spanish speakers were lower compared to scores obtained from the original English version. However, when the vocabulary items were matched for their frequency of occurrence in the original and target language and matched for meaning, test scores obtained from Spanish and English speakers were equivalent.

Similarly, across different languages, the same general category may have different prototypical members, and different words may be associated with each language for the same situation. These contextual variations make translated vocabulary tests particularly vulnerable to imbalance. When Peña, Bedore, and Zlatic-Giunta (in press) asked bilingual four- to six-year-olds to give examples of animals, the children's three most frequent English responses were "elephant," "lion," and "dog," while in Spanish they used "caballo" (horse), "elefante" (elephant), and "tigre" (tiger) in these orders.

In addition to vocabulary differences, grammatical structure also affects the validity of test translation practices. For example, nouns are marked by gender in Spanish, but not English. An English test translated to Spanish will miss aspects of Spanish, such as gender marking, that are not present in the English language. Furthermore, in Spanish, subject information is frequently carried in the verb, resulting in more complex verbs and less salient pronouns as compared to English. In English language assessment, pronoun omission is a hallmark of language impairment, yet this would not be true for Spanish. Thus, translated language tests may target inappropriate features for the target language, resulting in inaccurate assessment of language ability.

Problems Comparing Bilinguals and Monolinguals

Bilingual school children generally fall into the category of circumstantial bilinguals. That is, their circumstances (often a Spanish-speaking home and an English-speaking or bilingual school) require them to use two languages. These different environments typically require different language content. The home environment likely promotes discussions of common family activities, such as cooking or trips to the store, while more academic topics, such as colors, numbers, and shapes, are highlighted in the school environment. Bilingual children thus develop different vocabulary content for each language. From a testing perspective, this can result in underestimation of concept knowledge.

For example, Sattler and Altes (1984) examined typically developing three- to six-year-old bilingual Latino children's scores on the Peabody Picture Vocabulary Test-Revised and the McCarthy Perceptual Performance Scale. They found that the PPVT-R, whether administered in English or Spanish, yielded scores far below those of the norms, while all of the children were estimated to have normal intelligence based on their McCarthy scores.

A number of studies in the area of vocabulary acquisition illustrate that in early development, bilinguals learn unique words across their two languages, rather than learning two words (one in each language) for each concept. Pearson, Fernández, and Oller (1992) found that young bilinguals (8 to

30 months) often produced words for different concepts in each language, with few concepts labeled in both languages. Similarly Peña, Bedore, and Zlatic (in press) found that in a category generation task, bilingual children (ages 4 to 6 years) produced more unique words across Spanish and English than overlapping words.

When monolinguals and bilinguals are compared on measures of vocabulary, differences become more apparent. Pearson, Fernández, and Oller (1993) used the Spanish and English versions of the MacArthur Communicative Development Inventory (1989) to estimate bilingual toddler's vocabularies. They found that when compared to monolingual norms in either language, their scores were low. However, when they compared the total number of unique words they produced across the two languages, their scores were more comparable to the monolingual norms.

Another example of findings of differential performance between monolinguals and bilinguals is with the Test de Vocabulario en Imágenes Peabody: Adaptación Hispanoamericana (TVIP-H). This version of the Peabody Picture Vocabulary Test (PPVT) was normed on monolingual Spanish speakers outside of the U.S. mainland and then tested with bilingual Hispanics on the U.S. mainland. Bilinguals' scores were lower than those of the monolinguals (Dunn, 1988). The differences between monolinguals and bilinguals increased with age and coincided with schooling in English. Similarly, Umbel, Pearson, Fernandez, and Oller (1992) used the PPVT-Revised and the complementary Spanish version, the TVIP-H, to compare the receptive vocabularies of bilingual children ages 6 through 8 who were exposed to both Spanish and English in the home. On average, children responded correctly to 67% of the items in their age range in both languages, but another 8% to 12% were known only in one of their two languages. Administration of this test in only one language--even the "dominant" language--would have led to an underestimation of vocabulary knowledge.

Conceptual scoring (Pearson, Fernández, & Oller, 1993) has been proposed as a more meaningful measure of the bilingual's conceptual knowledge. The system, which entails counting the concepts demonstrated (either through constructed or selected responses) in both languages and correcting for concepts shared in the two languages, results in a more valid representation of a bilingual child's knowledge of concepts. The English/Spanish Bilingual Verbal Ability Tests (BVAT) (Cummins, Munoz-Sandoval, Alvarado, & Ruef, 1998) is based on this method.

Implications and Future Directions

The psychometric properties of a test, including item difficulty, item discrimination, reliability, and validity, do not automatically translate from one language to another, nor do they remain the same when a test is administered to a different audience than intended. Language tests can be improved if test developers:

- Ensure that concepts and linguistic features are appropriately represented for each language.
- Use conceptual scoring systems to eliminate understimation of ability.
- Select an appropriate mix of item types to gain the maximal amount of information about language ability in each language (e.g., an English grammar test may contain more emphasis on pronouns, while a Spanish grammar test might include more items related to gender and number agreement).
- Consider the frequency of occurrence of the words.

An important long-term goal is to better understand the development of language skills in bilinguals in order to develop language tests designed for, and normed on, bilinguals.

References

Cummins, J., Muñoz-Sandoval, A.F., Alvarado, C.G., & Ruef, M.L. (1998). *The Bilingual Verbal Ability Tests*. Itasca, IL: Riverside Publishing.

Dunn, L.H. (1988). *Bilingual Hispanic Children on the U.S. Mainland. A Review of Research on Their Cognitive, Linguistic, and Scholastic Development*. Honolulu, HI: Dunn Educational Services.

Figueroa, R. (1989). Psychological testing of linguistic-minority students: Knowledge gaps and regulations. *Exceptional Children, 56*, 145-148.

Pearson, B.Z., Fernandez, M.C., & Oller, D.K. (1992). Measuring bilingual children's receptive vocabularies. *Child Development, 63*, 1012-1221.

Pearson, B.Z., Fernandez, M.C., & Oller, D.K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning, 43*, 93-120.

Peña, E.D., Bedore, L.M., & Zlatic-Giunta, R. (in press). Development of categorization in young bilingual children. *Journal of Speech, Language, and Hearing Research*.

Restrepo, M.A., & Silverman, S.W. (2001). Validity of the Spanish Preschool Language Scale-3 for use with bilingual children. *American Journal of Speech-Language Pathology, 10*, 382-393.

Sattler, J.M., & Altes, L.M. (1984). Performance of bilingual and monolingual Hispanic children on the Peabody Picture Vocabulary Test-Revised and the McCarthy Perceptual Performance Scale. *Psychology in the Schools, 21*, 313-316.

Tamayo, J. (1987). Frequency of use as a measure of word difficulty in bilingual vocabulary test construction and translation. *Educational and Psychological Measurement, 47*, 893-902.

Umbel, V. M., Pearson, B.Z., Fernández, M.C., & Oller, D.K. (1992). Measuring bilingual children"s receptive vocabularies. *Child Development, 63*, 1012-1020.

Valdés, G., & Figueroa, R.A. (1994). *Bilingualism and Testing: A Special Case of Bias*. Norwood, NJ: Ablex.

BEST COPY AVAILABLE

This publication was prepared with funding from the Office of Educational Research and Improvement, U.S. Department of Education, under contract ED99CO0032. The opinions expressed in this report do not necessarily reflect the positions or policies of OERI or the U.S. Department of Education. **Permission is granted to copy and distribute this ERIC/AE Digest.**

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
X	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").